REMARKS

This amendment is in response to the Examiner's Office Action dated 1/3/2006. Reconsideration of this application is respectfully requested in view of the foregoing amendment and the remarks that follow.

STATUS OF CLAIMS

1. Claims 1-5, 7-11, and 17-19 are pending.
2. Claims 1-5, 7-11, and 17-19 stand rejected under 35 U.S.C. § 102(e) as being anticipated by Kraft et al. (6,516,312).

OVERVIEW OF CLAIMED INVENTION

The presently claimed invention allows a web crawler to accurately mimic real users, by relying on past user accesses to the Web sites to be crawled. This approach results in a web crawler capable of automatically accessing all the content that a real user would have access to.

In one embodiment, the present invention enumerates parameter combinations for automated access to World Wide Web content that mimics a real user access. Parameter combinations are based on input values that the real user has provided to input fields of a World Wide Web site. Parameter combinations are selected in a manner such that automated access patterns are "equivalent" real user access patterns. A log file maintains at least one set of parameters corresponding to a specific instance of real user interactions with a World Wide Web site. This log file is then analyzed to enumerate possible parameter combinations for achieving automated

access to the World Wide Web content "semantically equivalent" in nature to real user access patterns.

In another embodiment, a method of determining entries for input to an HTML form in pursuit of automated accesses to content contained in a Web database, is provided. Real user entries provided to an HTML form are logged and analyzed to enumerate combinations of entries for automatically populating an HTML file and the subsequent automated accesses to web content.

In the Claims

35 U.S.C. §102(e)

Applicants will set out many specific arguments below to distinguish over the Kraft reference used in the rejection. However, it is important for the examiner to recognize a very basic distinction between Kraft and the present invention. One distinction that limits Kraft from employing the claimed steps – Kraft generates all secondary searches from a "Result Set keyword", not the presently claimed "Query Log". When the examiner is pointing to figures 6A and 6B he is not pointing to query logs at all, only search result sets generated by queries (title of 6A is SEARCH RESULTS). Additional proof of this characterization can be found throughout Kraft. For example, col.1, lines 12-19 cite "....this invention pertains to a computer software product for dynamically associating keywords encountered in abstracts or summaries of a search result set....". Kraft uses the search results to match keywords with existing "dictionary terms" in

a local database to give additional info to the user. Claim 1, specifically claims a "query log"; claim 7 specifically claims "synthesis of entries"; and claim 17 specifically claims "a log containing real user entries". None of these claims is directed to an evaluation of search results as is taught by Kraft. Without a teaching of using a query log, Kraft cannot even satisfy the minimum required claim elements.

Applicants wish to emphasize that both the pending patent application and the primary reference (Kraft et al.) are commonly assigned and, at the time the claimed invention was made, were both subject to an obligation to be assigned to IBM. It will be shown below that the Kraft reference does not provide many of the elements of the claims and therefore cannot be properly rejected under 35 U.S.C. §102(e). A shift to a 35 U.S.C. §103 rejection would result in disqualification of this reference as prior art.

The examiner has rejected claims 1-5, 7-11, and 17-19 under 35 U.S.C. §102(e) as being anticipated by Kraft et al. (USP 6,516,312). To be properly rejected under 35 U.S.C. §102, each and every element of claims must be disclosed in a single cited reference. The applicants, however, contend that the presently claimed invention cannot be anticipated in view of the '312 reference.

The Kraft et al reference (hereafter Kraft) is primarily cited for its provision of new search queries generated from a domain-specific user query that was previously, dynamically associated with keywords. The Kraft reference teaches away from the present invention by generating a new search result from a set of previously prepared abstracts and by providing additional, supplemental information to each user query. Since a search engine repository is updated with

this additional information, subsequent executions of the same user query will not, and are not intended to generate the same or equivalent search results, but rather provide new, different information to the user.

With regard to independent claims 1, 7, and 17, the examiner has cited figures 6A and 6B of Kraft to equate to a log file containing queries. First, figures 6A and 6B are not log files, but rather search result listings (see title of figure – SEARCH RESULTS). As made clear above, figures 6A and 6B are simply search result sets with keywords noted. These keywords are matched to a local dictionary of terms (see col. 8, lines 35-38, etc.) – domain-specific dictionary 110.

By contrast, the present invention discloses an ordered set of parameters in a log file chosen such that automated access (as opposed to Krafts' user selecting of highlighted keywords for more info; see col. 11, lines 29-30, which state "the user, desiring to learn more about a desired term RMI, selects this term…", emphasis added) to the same WWW content as would be accessed manually, by a real user, is provided. Each parameter stored in a log file of the present invention is comprised of a name and associated value, specifically, an input field name in a WWW form and an associated input value to this field. In other words, the presently claimed invention seeks to reverse engineer a manual access of web content by automatically answering a question (i.e. input field name) presented by a web site with an answer (i.e. input value) that is based on a stored set of user responses (i.e. parameters values) to the same question (i.e. parameter name) presented by the same WWW form. A combination, as specified by the present invention, is a set of parameters that are individually input to a web form, whereas the combination disclosed by Kraft is number of distinct URLs and keywords combined to create a single query string.

Furthermore, Kraft teaches the automatic provision of a new search result to a browser for execution or to a web crawler for traversal. Therefore, applicants contend that abstracts contained in the log file maintained in search service provider cited by the examiner cannot be used to determine parameter combinations, nor can they be used in attaining access to web content, wherein access is automated or otherwise.

With respect to dependent claims 2, 3, and 6, the examiner has cited figure 6A as illustrating parameters and ranking. First, figure 6A illustrates a listing of search results. No teaching has been provided by the examiner directed to ranking query entries themselves. To simply show ranking, which is well known, without a teaching of ranking the same elements, leaves the argument without merit. The examiner must not only show ranking, but also that the ranking provides the same purpose or function - ranking of entries to queries.

The examiner appears to have equated parameters disclosed and claimed in the present invention with a text string and arbitrary URLs containing the same text string. Such a text string, for example, "RMI" in figure 6A as cited by the examiner, is not a parameter but rather, a simple keyword. A parameter of the present invention requires, for example, both a name (e.g. "zip code") and an associated value (e.g. "95120") appropriate for the name. In order for a web crawler to gain automated access to certain web content, the present invention teaches the determination of a value for a name component of a given parameter requested by a web site, for example a value for a zip code. In essence, when a Web site presents the question "What is the zip code?", a crawler reads previous responses stored for this question, answer the question with a value or values, "95120", based on what it read.

Because keywords and URLs are not parameters (i.e. they are not input fields with appropriately specified input values), their combination cannot be used to appropriately fill out HTML forms having fields requiring input. Additionally, keywords cited in figure 6a of the Kraft reference are not parameters because there are no input fields requiring input values.

With regard to dependent claims 4, 10, 11, and 18 the examiner has cited figure 6A of Kraft as suggesting both limit and unlimited text entries with removal of stop words and stemming. First, the examiner cites that he has noted certain stop words "by", "and" and "the" not within the search results. This argument bears no weight as no teaching of removing stop words has been made. To simply state that a specific chosen few stop words are not present does not equate to a removal step. In fact, the examiner has specifically chosen to ignore included stop words such as "with, and "or". Clearly stop words have not been removed. Secondly, pointing to an abbreviated term such as "monthly publication and author's full name" does not equate to "stemming remaining words". In fact, the term "programmer" is not stemmed. Clearly, the remaining terms are not stemmed.

With regard to dependent claims 5, 9, and 19, figure 3 of the Kraft reference is also cited by the examiner as suggesting a proxy server used in the description of the present invention. A proxy server of the present invention refers to a computer or program that is transparent to a client; a client does not see or know that a proxy server exists. Instead, a client sees web content produced by a web server to which the client is connected. A proxy server records communication between a web server and client silently and transparently (i.e. without requiring a client to know of its existence). In contrast, the search service provider pointed to by the examiner in the Kraft reference is, in fact, a web server directly providing content in response to

a client's request. It is implied, therefore, that a client is aware of the search service provider's existence by the fact that a client issues a request for content directly to the search service provider. Thus, a search service provider cannot be a proxy server for the following reasons: it provides content; a client is aware of a direct connection to it; and it does not intercept communications, transparently or otherwise.

As per claim 8, the examiner has equated storing annotated abstracts in a local database with maintaining a log file. However, no explicit recitation of a log file exists with Kraft.

SUMMARY

As has been detailed above, none of the references, cited or applied, provide for the specific claimed details of applicants' presently claimed invention, nor renders them obvious. It is believed that this case is in condition for allowance and reconsideration thereof and early issuance is respectfully requested.

As this amendment has been timely filed within the set period of response, no petition for extension of time or associated fee is required. However, the Commissioner is hereby authorized to charge any deficiencies in the fees provided to Deposit Account No. 09-0441.

If it is felt that an interview would expedite prosecution of this application, please do not hesitate to contact applicants' representative at the below number.

Respectfully submitted,

Randy W. Lacasse
Registration No. 34,368

1725 Duke Street
Suite 650
Alexandria, Virginia 22314
(703) 838-7683
April 3, 2006